ANALYSIS OF STATIONARY TIME SERIES

by

GAIL EUGENE BACHMAN

B. A., University of Wichita, 1959

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1963

Approved by:

A. M. Feyerherm

Major Professor

TABLE OF CONTENTS

BACKGROUND OF THE TIME SERIES PROBLEM

The usual model in least squares regression analysis is

$$X_t=\beta_0+\sum_{i=1}^{r}\beta_i Z_{it}+\epsilon_t,\ t=0,1,\ldots,n-1,$$

where the Z's are assumed fixed in repeated sampling and the ϵ's are independently distributed with mean zero and variance σ^2. In analysis of variance data the Z's may be merely dummy variates with values 0 or 1. When tests of significance or confidence limits are desired, normality of the ϵ's is also assumed.

In the sciences many problems occur in which a process produces what may be considered a family of random variables such that there is a value of x_t for each value of t in some interval T. The experimenter wishes to investigate the nature of the response curve over the interval T. One of the major difficulties in the application of traditional statistical methods to these time series data is the possible absence of independence of successive observations. If the ϵ's are not independent the assumptions necessary for using ordinary least squares estimation theory are violated. It is the correlation of the ϵ's and not of the X's which is to be avoided. Attitudes of research workers toward regression analysis of time series have varied between widely separated extremes. Until the middle 1920's, many researchers were completely unaware of the problems connected with the sampling of time series. Following the appearance of articles such as Yule's (1926) on "non-sense correlations," it was maintained that existing methods simply did not apply to time series and that

reputable statisticians should leave time series alone. Koopmans, Wold and others clarified the sampling significance of regression analysis based on time series in the late 1930's. Considerable work followed on the problem of testing for the existence of correlation of the errors but all too little on the more important problem of the best estimation procedure when the correlations do exist. Results are still somewhat lacking in this latter area, but several estimation procedures have been proposed since 1950, some by social and natural scientists, and others by physical scientists and engineers. The method of spectrum analysis is most prominent in the latter category. This paper will deal primarily with those methods generally used in the social and biological sciences.

NATURE OF THE TIME SERIES PROBLEM

Stationary Time Series

The discussion of time series is usually confined to what are called stationary processes or stationary time series. There are two important types of stationarity. A process is called strictly stationary if the distribution of the set

$$(x_{t_1}, \ldots, x_{t_n})$$

of random variables from $(x_t : t \in T)$ is the same as that of the set

$$(x_{t_1+h}, \ldots, x_{t_n+h})$$

for every n, t_1, $t_2, \ldots, t_n$ and h. This roughly means that the time series is without trends, not only in the mean values of the

x_t but also in their variances. Most of the studies in time series do not require the assumption of strict stationarity but are based on the weaker assumptions that: (1) $E(x_t)$ is a constant for all t which may be taken as zero, and (2) the distributions above have the same covariance matrix for all h. A time series $(x_t:t\in T)$ is said to be weakly stationary if it satisfies these two conditions. Hence the covariance matrix depends only upon the time differences

$$t_2-t_1,\ t_3-t_2,\ldots,\ t_n-t_{n-1},$$

and the covariance of x_{t+h} and x_t is a function of h only. If $E(x_t)$ is taken to be zero, then $E(x_t x_{t+h})=\gamma_h$. The covariance γ_h is usually called the autocovariance between x_t and x_{t+h}, and ρ_h is called the autocorrelation function of lag h.

For some time series $(y_t:t\in T)$ the model will have the form

$$y_t=m_t+x_t$$

where m_t is a constant for each t and $(x_t:t\in T)$ is a stationary time series with $E(x_t)=0$. Since $E(y_t)=m_t$, the covariance function of $(y_t:t\in T)$

$$\gamma_h=E\left[(y_{t+h}-m_{t+h})(y_t-m_t)\right]=E(x_{t+h}x_t)$$

is identical with that of $(x_t:t\in T)$. Estimating m_t and γ_h from a finite number (n) of observations taken from the time series is one of the problems of time series analysis.

Models

There are a number of models which may be used in analyzing time series. If it is assumed that the data follow an underlying

systematic scheme with random fluctuations superimposed, the methods of harmonic analysis and periodogram analysis may be used to determine the nature of the systematic component for functions with regular periods. In cases where the periods are known, for example, seasonal variation studies in economics, harmonic analysis is used to determine the amplitudes. Where the periods are regular but unknown, periodogram analysis can be used to seek out the hidden periodicities. If the systematic movement is oscillatory with irregular periods, the variables Z_{it} in the regular regression model may become t^i, and a polynomial form used to locally describe the systematic component.

In other cases, the assumed model may involve lagged values of X as predictors. An example of this autoregressive model might be

$$X_t=\beta_0+\sum_{i=1}^{r}\beta_i X_{t-i}+\epsilon_t.$$

Or finally, a combined regression model could be used with lagged X's, present Z's and lagged Z's as predictors.

The choice between models is very difficult. It may happen that one model fits well and the others rather poorly. For short series it is usually impossible to determine whether this phenomenon is due to the choice of the model or to the particularities of the sample analyzed. Particularly for the autoregressive schemes, tests for goodness of fit are not well developed.

Testing a Series for Autocorrelation

Suppose $(x_1, \ldots, x_n)$ is a sample from a normal time series

(x_t) and the hypothesis to be tested is that the time series $x_1, \ldots, x_n$ are independent random variables having identical normal distributions $N(\mu, \sigma^2)$. The term white noise is often used in reference to such independent random variables.

R. L. Anderson (1942) proposed a criterion for testing this hypothesis with the ratio

$$R_1 = c_1'/c_0'$$

where $\sigma^2 c_h' = \sum_{\xi=1}^{n} (x_\xi - \bar{x})(x_{\xi+h} - \bar{x})$, $h=0,1$, and $x_{n+1}=x_1$. Use of the relation $x_{n+1}=x_1$, as opposed to running the summation from t=1 to n-1, is somewhat arbitrary, but it simplifies the distribution theory of R_1. If a sample is from a white noise, then with n large, R_1 will tend to have values near 0. If the sample is not from a white noise, then R_1 will tend to have values away from 0.

Anderson derived the sampling distribution for R_1 and has prepared tables for $Pr[R_1 > R_1(\alpha)] = \alpha$, for $\alpha=0.99$, 0.95, 0.05, 0.01, and n=5(1)15(5)75. Values of R_L for lag other than 1 may be tested using the table for R_1, since for large samples R_L is approximately distributed like R_1. For large n, Anderson also showed that R_L is approximately normally distributed with mean $-1/(n-1)$ and variance $(n-2)/(n-1)^2$.

Koopmans (1942) examined R_1 as an estimate of ρ in the simple autoregressive model

$$X_t = \rho X_{t-1} + \epsilon_t.$$

The circular definition of R_1 was not satisfactory if the alternative hypothesis specified this form. Von Neumann (1941) had earlier obtained the distribution of

$$\frac{\delta^2}{s^2} = \frac{1}{n-1}\sum_{t=1}^{n-1}(x_{t+1}-x_t)^2 \Big/ \frac{1}{n}\sum_{t=1}^{n}(x_t-\bar{x})^2.$$

Hart (1942) tabulated the probabilities by use of a series approximation. T. W. Anderson (1954) then showed that

$$R_c = \frac{1/2\left[(x_1-\bar{x})^2 + (x_n-\bar{x})^2\right] + \sum_{t=1}^{n-1}(x_t-\bar{x})(x_{t+1}-\bar{x})}{\sum_{t=1}^{n}(x_t-\bar{x})^2}$$

had greater power than R_1 in testing the hypothesis that $\rho=0$ for Koopmans' model and that no uniformly most powerful test exists for such a hypothesis. Since

$$\frac{\delta^2}{s^2} = 2n(1-R_c)/n-1,$$

Anderson was able to transform Hart's significance levels into significance levels for R_c.

A non-parametric test of great simplicity is due to Wald and Wolfowitz (1943), but it is also circularly defined and somewhat limited in use. Many other papers on testing for autocorrelation have appeared; those mentioned here are probably the most significant.

Testing for Autocorrelation in Residuals

In the previous section on testing for autocorrelation the process (x_t) considered had mean value zero and, if necessary, the mean correction was applied. The general class of time processes with which the experimenter is usually concerned will need to be reduced to stationary form by simple subtraction of a time

dependent mean. Such a preliminary treatment of data will nearly always be necessary before methods of stationary time series can be applied.

Consider the linear regression of a variable y_t upon k regressor variables $Z_{1t}, Z_{2t}, \ldots, Z_{kt}$. Then y_t is regarded as being generated by a relation of the form

$$y_t = \beta_0 + \sum_{i=1}^{k} \beta_i Z_{it} + x_t$$

where x_t is generated by a stationary process. The Z_{it} are considered fixed and independent of the x_t and inferences made conditionally upon the fixing of the Z_{it} at their observed values.

It has been shown, as shall be more fully illustrated in a later section, that the departure of the process generating x_t from a process generating independent random variables may effect both the efficiency of the least squares methods and the validity of the usual tests of significance. Lacking any precise prior knowledge as to the nature of the data, a reasonable procedure may be to carry out an initial regression based on the assumption that the x_t's are white noise. The x_t can then be tested for mutual independence. The fact that the β's are estimated invalidates the use of the methods of detecting autocorrelation presented in the previous section.

A small sample test of the null hypothesis that the x_t are independent and normal with zero mean is due to Durbin and Watson (1950, 1951). Let the n successive least squares residuals be $V_1, V_2, \ldots, V_n$. A modification of the von Neumann statistic

$$d = \frac{\sum_{i=1}^{n-1} (V_{i+1}-V_i)^2}{\sum_{i=1}^{n} V_i^2}$$

is used to test for the existence of autocorrelation in the residuals. It will be noted that

$$\frac{\delta^2}{s^2} = \frac{nd}{n-1} \text{, and } d=2(1-R_c)$$

but since the original von Neumann and T. W. Anderson statistics did not refer to the residuals from a regression analysis, tables for those statistics cannot be used here. An exact distribution for d cannot be evaluated, but upper and lower significance bounds, d_u and d_L, could be computed. This was done by Durbin and Watson for 5%, 2.5%, and 1% one-tailed tests, for n=15(1)40 (5)100 and for k=1(1)5. It should be noted that d_u and d_L will diverge as k increases and also as n increases.

In most cases the experimenter desires a test of the null hypothesis against the alternative of positive correlation. The expected value of d will be small when the null hypothesis is false, so if the computed value of d is less than the tabulated value the null hypothesis is rejected. If the alternative hypothesis was negative correlation, d would be expected to be near 4. In this case $d'=4-d$ is considered and tested against the tabulated value as above. Durbin and Watson present alternative approximation methods for use when n is greater than 40.

Moran (1950) presented an exact test for the residuals from regression when only one predictor is used. He used the first two

autocorrelation coefficients of the residuals, defined in a circular fashion, and showed that the expected value of the autocorrelation coefficient of the residuals r_1 is

$$E(r_1) = \frac{-(1+R_1)}{N-2}$$

and that

$$E(r_1^2) = \frac{N+1}{N^2} + \frac{2R_1+3R_1^2-2R_2}{N(N-2)}$$

Finally Moran shows that for large samples the quantity

$$\frac{r_1-E(r_1)}{\sigma_{r_1}}$$

is normally distributed with mean zero and variance one.

The Correlogram

A useful tool in the analysis of time series, first proposed by Wold, is called the correlogram. Wold (1953) indicated that the choice of possible models used to explain stationary time series data depended upon the relationship of successive true autocorrelation coefficients ρ_L. The sample values R_L are usually displayed graphically as in Fig. 1.

Three possible forms of the correlogram are readily apparent. First the curve may be strictly periodic with repeated non-damped cycles. This suggests the use of harmonic analysis. Secondly, the curve may be damped but with $|\rho|$ greater than zero. This type of curve may be generated by a linear autoregressive model. The

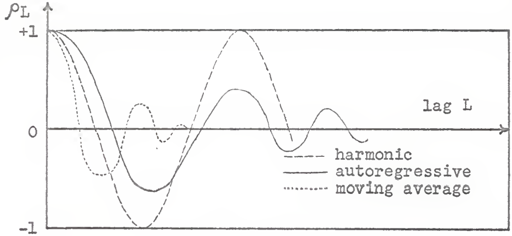

Fig. 1. Correlogram.

third alternative is a damped correlogram with ρ_L equal to 0 for some L greater than m. Wold suggests the use of moving averages to transform the data to non-autocorrelated observations.

The extent to which the fine structure of a correlogram can be interpreted seems limited and it appears best to concentrate on certain features such as pronounced oscillations and the speed with which the R_L converge to zero. Bartlett (1946) has shown that successive autocorrelation coefficients tend to be autocorrelated and hence caution should be used in determining the model from the correlogram. Especially with relatively short time series, the empirical correlogram may depend more upon the properties of the sample than upon the population, but it still is a valuable tool in selecting a suitable model.

To illustrate these concepts, precipitation data for Manhattan, Kansas, was obtained and the probability of a dry day was calculated for each day of the year. These probabilities are listed in the Appendix. Now if it is assumed that the observed probabilities are the result of random variation superimposed on

a systematic model, the correlogram for the 365 observations may give an indication of the appropriate model.

The autocorrelation coefficient (circular-definition) for possible lags are listed in Table 1 and plotted in Fig. 2. Although care must be taken in drawing conclusions from the correlogram, the oscillation suggests that either an autoregressive model or a periodic model be used to estimate the systematic component of the data. It will be shown later that harmonic analysis used to fit a periodic function gives errors which appear random.

Table 1. Coefficients of Autocorrelation for Probabilities of Dry Days in Manhattan, Kansas.

Lag	R_L	Lag	R_L	Lag	R_L	Lag	R_L
1	.6711	95	-.0899	190	-.5405	285	.0825
5	.6313	100	-.1065	195	-.5017	290	.1317
10	.6246	105	-.1557	200	-.5143	295	.1811
15	.5932	110	-.1437	205	-.4711	300	.1860
20	.5712	115	-.2766	210	-.4543	305	.2565
25	.5189	120	-.2720	215	-.4534	310	.2893
30	.5157	125	-.2869	220	-.4452	315	.3421
35	.4645	130	-.3342	225	-.4035	320	.3479
40	.4124	135	-.3775	230	-.3775	325	.4123
45	.3479	140	-.4035	235	-.3342	330	.4645
50	.3421	145	-.4452	240	-.2869	335	.5156
55	.2893	150	-.4533	245	-.2721	340	.5188
60	.2566	155	-.4543	250	-.2766	345	.5712
65	.1860	160	-.4710	255	-.1437	350	.5932
70	.1811	165	-.5142	260	-.1558	355	.6246
75	.1317	170	-.5017	265	-.1065	360	.6313
80	.0825	175	-.5405	270	-.0899		
85	.0221	180	-.5310	275	-.0295		
90	-.0295	185	-.5310	280	.0221		

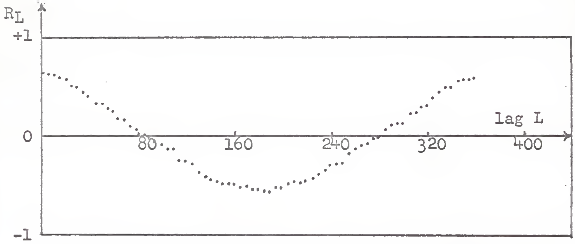

Fig. 2. Correlogram for Probabilities of Dry Days.

THE AUTOREGRESSIVE MODEL

In many fields of study the time series phenomena may be represented by a regression model of the form

$$x_t+\alpha_1 x_{t-1}+\dots+\alpha_p x_{t-p}=\beta_1 Z_{1t}+\dots+\beta_q Z_{qt}+\epsilon_t,$$

$$t=0,1,\dots,n-1, \qquad (1)$$

where ϵ_t is a series of independently and identically distributed random variables with mean zero and variance σ^2. This is a generalization of both the ordinary regression model

$$x_t=\beta_1 Z_{1t}+\dots+\beta_q Z_{qt}+\epsilon_t$$

and of the autoregression model

$$x_t+\alpha_1 x_{t-1}+\dots+\alpha_p x_{t-p}+\alpha_0=\epsilon_t \,.$$

Values of x_0, x_{-1}, ..., x_{-p+1} are usually regarded as given numbers, or if they are considered as random variables, inferences are made conditionally on those quantities held fixed.

Since the coefficients in the normal equations are random variables for model (1), rather than constants as for ordinary regression models, difficulties arise in finding the sampling distributions of the least-squares estimators.

Model with Lagged Dependent Variable

Mann and Wald (1943) studied the autoregression model $x_t+\alpha_1 x_{t-1}+\cdots+\alpha_p x_{t-p}+\alpha_0=\epsilon_t$ and showed that ordinary least squares theory is valid asymptotically.

A method presented by Durbin (1960) examines the properties of estimators for the model containing lagged x's. He considered the simplest cases of the regular regression model, namely,

$$x_t=\beta Z_t+\epsilon_t,\ t=1,\ldots,n$$

where $Z_1,\ldots, Z_n$ are constants, and

$$x_t+\alpha x_{t-1}=\epsilon_t,\ t=1,\ldots,n$$

where x_0 is constant. Both cases take ϵ_t to be independently and identically distributed with mean zero and variance σ^2.

Application of least squares gives estimates for β and α of the form

$$b=\sum_{t=1}^{n} x_t Z_t/\sum_{t=1}^{n} Z_t^2 \quad\text{and}\quad a=-\sum_{t=1}^{n} x_t x_{t-1}/\sum_{t=1}^{n} x_{t-1}^2$$

respectively. The estimate b is the minimum-variance unbiased estimator for β, but a is biased and its small sample properties do not follow directly from any classical theory. The difference arises from the fact that while b is a linear function of the x's and relatively easy to handle, a is a ratio of quadratic forms.

In developing a reasonable optimality criterion for estimating

α, Durbin considers the estimating equation

$$a\sum_{t=1}^{n} x_{t-1}^{2}+\sum_{t=1}^{n} x_t x_{t-1}=0$$

from which a is derived. If a is replaced by α then

$$E(\alpha\sum_{t=1}^{n} x_{t-1}^{2}+\sum_{t=1}^{n} x_t x_{t-1})=0.$$

The linear equation in a is called an unbiased estimating equation in accordance with a definition by Durbin:

> Suppose that the estimator a of a parameter α is given by the linear equation
>
> $$T_1 a+T_2=0, \tag{2}$$
>
> where T_1 and T_2 are functions of the observations such that T_2/T_1 is independent of unknown parameters, and where
>
> $$E(T_1\alpha+T_2)=0. \tag{3}$$
>
> Then equation (2) is called an unbiased linear estimating equation.

Linear in this case means linear in a, not linear in the observations. The quantity T_1 is assumed to be non-zero. If $T\equiv 1$, this definition includes the ordinary notion of an unbiased estimator.

A second definition is necessary in requiring the analogue of minimum variance of an unbiased estimator. Since (2) may be multiplied through by an arbitrary constant without affecting the value of a, requiring $T_1\alpha+T_2$ to have minimum variance is not enough. To take care of this situation, the equation was standardized by dividing through by $E(T_1)$.

If $t_1=T_1/E(T_1)$ and $t_2=T_2/E(T_1)$ then Durbin's second definition is:

> Suppose that $t_1 a+t_2=0$ is an unbiased linear estimating

equation where $E(t_1)=1$ and

$$V(t_1\alpha+t_2)\leq V(t_1'\alpha+t_2'), \tag{3}$$

for all other unbiased linear estimating equations $t_1'a+t_2'=0$ having $E(t_1')=1$. Then $t_1a+t_2=0$ is called a best unbiased linear estimating equation.

The notion of a minimum variance unbiased estimator is included in this definition, for if $t_1\equiv t_1'\equiv 1$ and (2) is satisfied, a is a minimum variance unbiased estimator of α.

Now a lower bound for the variance of $t_1\alpha+t_2$ is derived. Let $T_1a+T_2=0$ be an unbiased estimating equation where T_1 and T_2 depend only on the observations. If the sample density is $\phi(x_1,\ldots, x_n;\alpha)$, then from (3)

$$\int_R (T_1\alpha+T_2)\ \phi\ dx=0$$

where $\int_R$ denotes the multiple integral and dx stands for $dx_1,\ldots, dx_n$. If the conditions for differentiating under the integral sign are satisfied, the differentiation with respect to α gives

$$\int_R t_1\phi\ dx + \int_R (t_1\alpha+t_2)(\partial\phi/\partial\alpha)dx=0.$$

Since $E(t_1)=\int_R t_1\phi\ dx=1$ and $\partial\phi/\partial\alpha=\phi\ \partial\log\phi/\partial\alpha$, one may write

$$\int_R (t_1\alpha+t_2)(\partial\log\phi/\partial\alpha)\phi dx=-1.$$

By Schwarz's inequality

$$E(t_1\alpha+t_2)^2\ E(\partial\log\phi/\partial\alpha)^2\geq\left[\int_R(t_1\alpha+t_2)(\partial\log\phi/\partial\alpha)\phi dx\right]^2=1,$$

so that finally

$$V(t_1\alpha+t_2)\geq\frac{1}{E(\partial\log\phi/\partial\alpha)^2}=-\frac{1}{E(\partial^2\log\phi/\partial\alpha^2)}. \tag{4}$$

If $t_1 \equiv 1$ then $-t_2$ is an unbiased estimator and (4) becomes the Cramér-Rao inequality

$$E(t_2+\alpha)^2 \geqslant \frac{1}{E(\partial \log \phi/\partial\alpha)^2}$$

for the lower bound on the variance of an unbiased estimator.

Quantity t_1 may not be identically equal to one but may converge stochastically to one as $n \longrightarrow \infty$. In this case let δ_n be a function of n such that

$$E(\partial \log \phi/\partial\alpha)^2 = 0(\delta_n^2).$$

It follows that since $t_1(a-\alpha)=-(t_1\alpha+t_2)$, the limiting minimum variance is that found above. Thus the asymptotic distribution of $\delta_n(a-\alpha)$, if it exists, has mean zero and the limiting minimum variance is

$$\lim_{n\longrightarrow\infty} \frac{\delta_n^2}{E(\partial \log \phi/\partial\alpha)^2}$$

For single parameter problems δ_n is usually equal to $\sqrt{n}$.

As an example, the errors $\epsilon_1, \ldots, \epsilon_n$ for the autoregressive model can be assumed normally distributed with unit variance. The density function is then

$$= \frac{1}{(2\pi)^{n/2}} \exp\left[-1/2 \sum_{t=1}^{n} (x_t+\alpha x_{t-1})^2\right].$$

Application of the method of maximum likelihood gives the linear estimating equation, namely,

$$a \sum_{t=1}^{n} x_{t-1}^2 + \sum_{t=1}^{n} x_t x_{t-1} = 0,$$

which is unbiased according to the previous definition.

Differentiation leads to

$$-\partial^2 \log \phi / \partial\alpha^2 = \sum_{t=1}^{n} x_{t-1}^2 ,$$

so from the previous derivation, with $T_1 = \sum_{t=1}^{n} x_{t-1}^2$ and $T_2 = \sum_{t=1}^{n} x_t x_{t-1}$ one obtains

$$V(t_1\alpha + t_2) \geqslant \frac{1}{E(\sum_{t=1}^{n} x_{t-1}^2)} .$$

For this example this lower bound is actually attained as

$$t_1\alpha + t_2 = -\frac{\partial \log \phi}{\partial\alpha} \Big/ E(\sum_{t=1}^{n} x_{t-1}^2)$$

and

$$V(t_1\alpha + t_2) = \frac{E(\partial \log \phi/\partial\alpha)^2}{\left[E(\sum_{t=1}^{n} x_{t-1}^2)\right]^2} = \frac{1}{E(\sum_{t=1}^{n} x_{t-1}^2)} .$$

To consider the variance of the estimator a, asymptotic theory and the assumption $|\alpha| < 1$ are used. The expected value,

$$\frac{1}{n} E(\sum_{t=1}^{n} x_{t-1}^2) \longrightarrow \frac{1}{1-\alpha^2}$$

as n becomes large, so that in the limit $\sqrt{n}(a-\alpha)$ will have zero mean and variance $1-\alpha^2$. Thus a is an asymptotically efficient estimator of α since $1-\alpha^2$ is the minimum variance possible.

Durbin extends this proof to multi-parameter problems and shows that, in general, the same properties hold. These results are important for the next section where the model has autocorrelated errors.

Hurwicz (1950) studied the small sample bias of the parameters in autoregressive models and indicated the serious proportions that the bias may take on.

Model with Autocorrelated Errors

For many situations, the appropriate model is

$$x_t=\beta_1 Z_{1t}+\ldots+\beta_q Z_{qt}+U_t, \quad (t=0,\ldots,n-1) \tag{5}$$

where (U_t) is a stationary autoregressive series given by

$$U_t+\alpha_1 U_{t-1}+\ldots+\alpha_p U_{t-p}=\epsilon_t, \quad (t=\ldots-1,0,1,\ldots) \tag{5a}$$

and where the Z's are given constants. The ϵ_t's are assumed to be independently and identically distributed with mean zero and variance σ^2. This model differs from the model of the previous section in that it does not contain lagged x's and has autocorrelated error terms.

The common assumption of independence of error terms may be violated in data such as a series of outputs of a production process. Cochrane and Orcutt (1949) have offered three reasons why the ϵ_t's in economic time models tend to be autocorrelated:

1. Use of incorrect functional form of the relationship.
2. Omitted variables are usually autocorrelated.
3. Errors of measurement are often autocorrelated.

They conducted some empirical sampling studies using generated autoregressive error processes with a given regression model. The series used were analogous in length to most available economic time series with approximately twenty observations. When least

squares regression was used to analyze the generated series, the results indicated:

1. Estimated autocorrelation of the residuals tended to be biased toward randomness.
2. The least squares estimates are not biased even though they are not the best estimates.
3. When the autocorrelation of the errors is high the variance of the least squares estimates is greatly increased.
4. Nearly optimum results can be achieved if the error term is only a rough approximation to a random series so even a simple transformation of the error term may be adequate.
5. If sample residuals are used to estimate the error variance, σ^2, this estimate will be too small when the errors are positively correlated.
6. Analyzing first differences is a good method for economic problems.

Champernowne (1948) showed similar results by theoretical work with this model.

Now returning to the original model, if the U_t, $U_{t-1},\ldots,U_{t-p}$ are expressed in terms of the x's and Z's using (5), the model (5a) becomes

$$x_t+\alpha_1 x_{t-1}+\ldots+\alpha_p x_{t-p}$$

$$=\beta_1 Z_{1t}+\ldots+\beta_q Z_{qt}+\alpha_1\beta_1 Z_{1,t-1}+\ldots+\alpha_p\beta_p x_{q,t-p}+\epsilon_t.$$

An investigation of the efficiencies and estimated variances of least squares estimates of regression coefficients for fixed Z's and tests of hypotheses concerning them when an incorrect transforming model is used has been carried out by Watson (1951). Various types of general solutions are presented: bounds on the bias of the estimated variance, lower bound on the efficiency of the estimates of regression coefficients and some bounds on the significance points of the t and F tests. Some special types of

incorrect transformations are also discussed. It was found that for what appeared to be only mildly inaccurate estimates, the true probabilities for 5% significance levels may be considerably different. Watson takes a rather pessimistic view of the use of transforming devices to remove the effect of autocorrelation in time series data.

Application of least squares to this equation will, in principle, lead to optimum estimates when the ϵ_t are normally distributed. These equations will be non-linear and hence, difficult to solve. Some sort of iterative procedure is required. Various methods have been suggested by Champernowne (1948), Cochrane and Orcutt (1949), Durbin (1960) and others, but these are computationally inefficient.

Fuller and Martin (1961) have suggested the simultaneous estimation of both the error sturcture and the model by least squares. This method appears much more promising for practical work. For the first order autoregressive error and a single Z variable this method is easily illustrated. The models are

$$x_t=\beta Z_t+U_t \tag{6}$$

and

$$U_t+\alpha U_{t-1}=\epsilon_t. \tag{7}$$

Substituting (6) and (6) lagged into (7) gives

$$x_t=\beta Z_t+\alpha\beta Z_{t-1}-\alpha x_{t-1}+\epsilon_t. \tag{8}$$

Estimation of α and β is now clearly a problem in non-linear estimation. If the equation is rewritten as

$$x_t=\theta_1 Z_t+\theta_2 Z_{t-1}+\theta_3 x_{t-1}+\epsilon_t$$

where $\theta_2=-\theta_1\theta_3$, the problem may now be viewed as a non-linear restriction upon the three parameters. Independence of the errors will now be a special case with $\alpha=0$.

This problem can now be handled by the modified Gauss-Newton iterative procedure. The problem becomes one of regression by expanding (8) in a Taylor's series about a point $P_o=(\alpha_o,\beta_o)$, where α_o and β_o are guessed values of the parameters. If only the first order terms are considered, then

$$x_t-x_{to}\doteq(Z_t+\alpha_o Z_{t-1})\Delta\beta+(\beta_o Z_{t-1}-x_{t-1})\Delta\alpha$$

where $\Delta\beta=\beta-\beta_o$ and $\Delta\alpha=\alpha-\alpha_o$. The corrections in the trial values, $\Delta\alpha$ and $\Delta\beta$, can be found by regressing (x_t-x_{to}) on $(Z_t+\alpha_o Z_{t-1})$ and and $(\beta_o Z_{t-1}-x_{t-1})$.

Hartley (1961) has shown that the residual sum of squares decreases in the Gauss direction, that is, that some $k>0$ exists such that the residual sum of squares associated with $P_{ok}=(\alpha_o+k\Delta\alpha,\beta_o+k\Delta\beta)$ is less than the residual sum of squares associated with P_o. It may happen that the full step results in an increase in the residual sum of squares. In order to assure a decrease in the residual it is necessary to compare the preceding residual with the computed residual sum of squares at the end of each iteration. If a decrease is recorded, $(\alpha_o+\Delta\alpha,\beta_o+\Delta\beta)$ are used as start values for the next iteration. If a decrease is not recorded, the start values are taken as $(\alpha_o+1/2\ \Delta\alpha,\beta_o+1/2\ \Delta\beta)$ and the residual sum of squares computed. If a decrease is not noted at this step, the residual sum associated with $(\alpha_o+1/4\ \Delta\alpha,\beta_o+1/4\ \Delta\beta)$ is found and

so on until the decrease occurs. The iteration is carried on until the $\Delta\alpha$ and $\Delta\beta$ satisfy a criterion of form

$$\frac{(\Delta\alpha_i)^2}{\text{var }(\hat{\alpha}_i)} \leq C.$$

The serious problem may be in locating an initial approximation (α_o, β_o) in the region of the absolute minimum of the residual. A preliminary grid over a wide range for α and β may be necessary to find a sufficiently close approximation. The absolute minimum and not a local minimum must be found.

If the Z_{it} are assumed to be bounded and the ϵ_t normally distributed the final set of estimates are maximum likelihood estimates possessing the properties of consistency and asymptotic normality. Large sample variances and covariances are estimated in the ordinary manner as the product of the elements, C_{ij}, of the inverse of the variance-covariance matrix at the final iteration and the estimated variance s^2. The variance is estimated by

$$s^2 = \frac{\sum_{t=0}^{n-1} (x_t - \hat{x}_t)^2}{n-r}$$

where r is the number of parameters estimated.

The exact nature of the correlation properties is of course unknown. A second order autoregressive scheme

$$U_t + \alpha_1 U_{t-1} + \alpha_2 U_{t-2} = \epsilon_t$$

could be assumed and the parameters estimated in a similar way.

Goodness-of-fit tests for autoregressive schemes have not been developed to a sufficient extent. It seems though that for most situations, a model only roughly approximating the true one will give the desired random error term.

PARAMETRIC TIME SERIES

The classical model which has been used widely in time-series analysis consists of two parts, a systematic part M_t, and a random element of error ϵ_t with mean zero and variance σ^2. If the observed item is x_t $(t=0,1,\ldots,n-1)$ the time series has the form

$$x_t=M_t+\epsilon_t \ .$$

The stochastic element ϵ_t is superimposed on the non-stochastic part M_t and the error at one time point does not affect a later observation. This model is not valid if the error elements are autocorrelated.

Different methods of analysis are appropriate for different assumptions about the nature of M_t. If the data indicate that M_t is a "smooth" function of time, that is, M_t is not highly irregular or periodic in form, a polynomial may be used to locally represent the data. The autocorrelations of M_t and M_{t+h} $(h=1,2,\ldots,n)$ should be positive, zero, or small negative numbers. A semi-empirical procedure known as the variate difference method is commonly used to estimate the degree of this polynomial.

When oscillatory and periodic movements are present in the data the function to be fitted must be of trigonometric form. This usually involves the use of Fourier analysis or some related

procedure.

Variate Difference Method

Suppose the time series x_t, $t=\ldots-1,0,+1,\ldots$, is known to be of the form

$$x_t=\sum_{p=0}^{k}\beta_p t^p +\epsilon_t=M_t+\epsilon_t$$

where $\beta_0,\beta_1,\ldots,\beta_k$ are unknown and where ϵ_t is a random element with variance σ^2.

For a sample $(x_1,\ldots, x_n)$, $n>k+1$, minimum variance estimators for the β's can be obtained by least squares and the variances of the estimators calculated by usual methods if the parameter k is known. However, k is usually unknown and must also be estimated.

A polynomial of degree p has the well-known property that its (p+1)th finite differences vanish. Tintner (1940) has used this property in developing the variate difference method for estimating the value of k for the given model.

Let y_t be the time series defined by the hth forward difference of the time series x_t. Since Δ is a linear operator

$$y_{h,t}=\Delta^h x_t=\Delta^h M_t+\Delta^h\epsilon_t.$$

By the advancing difference formula

$$\Delta^h\epsilon_t=\epsilon_{t+h}-\binom{h}{1}\epsilon_{t+h-1}+\binom{h}{2}\epsilon_{t+h-2}-\ldots+(-1)^h\epsilon_t$$

and

$$\Delta^h M_t=M_{t+h}-\binom{h}{1}M_{t+h-1}+\binom{h}{2} M_{t+h-2}-\ldots+(-1)^h M_t.$$

If one considers the sequence of samples $(x_1,\ldots,x_{n+h})$, $h=1, 2,\ldots$ and forms the ratios

$$Q_h=\sum_{\xi=1}^{n} y_{h,\xi}^2/\left[n\binom{2h}{h}\right],\ h=1,2,\ldots$$

it follows that

$$E(Q_h)=E\sum_{\xi=1}^{n}\ (\Delta^h M_\xi+\Delta^h\epsilon_\xi)^2/\left[n\binom{2h}{h}\right]$$

$$=\sum_{\xi=1}^{n}\left[(\Delta^h M_\xi)^2+E(\Delta^h\epsilon_\xi)^2/\left[n\binom{2h}{h}\right]\right],$$

since M_ξ is a nonrandom function of t and ϵ_ξ is a random element.

From the above expression for $\Delta^h\epsilon_t$ we find

$$E(\Delta^h\epsilon_\xi)^2=\sum_{p=0}^{h}\binom{h}{p}^2 E\left[(\epsilon_{\xi+h-p}^2)\right]=\sigma^2\sum_{p=0}^{h}\binom{h}{p}^2=\binom{2h}{h}\sigma^2.$$

The first term is always non-negative and will vanish for all $h\geqslant k+1$. Therefore $E(Q_h)=\sigma^2$ for $h\geqslant k+1$.

In a practical situation, the question then becomes: Which difference series sufficiently explains the non-random part of the time series so that all difference series of higher order are estimates of σ^2 and represent ϵ_t alone?

Under the assumption that element ϵ_t is normally distributed with mean zero and variance σ^2, a large sample test has been given by O. Anderson (1929) for testing the hypothesis that the variance of the difference series of order h is approximately equal to the variance of difference series h+1, i.e. $Q_h=Q_{h+1}$. For a sample $(x_1,\ldots, x_n)$, $n\gg k_o$ the estimates of the variances of the difference series are

$$Q_h = \sum_{\xi=1}^{n-h} y_{h,\xi}^2 \Big/ \left[(n-h)\binom{2h}{h}\right]$$

If the systematic part of M_t has been eliminated in the finite difference series of order h_o, then approximately

$$Q_{h_o} = Q_{h_o+1} = Q_{h_o+2} = \cdots .$$

In order to test the approximate equality of Q_h and Q_{h+1}, the standard error of $Q_{h+1} - Q_h$ is computed:

$$e_h = \frac{Q_h}{H_{hn}}, \; k=0,1,2,\ldots .$$

H_{hn} has been tabulated by Tintner (1940). An asymptotic formula can be used for large values of n and $h > 6$:

$$e_h^2 = \frac{(3h+1)Q_h^2\sqrt{2\pi h}}{2(2h+1)^3(n-h-1)}, \; h=6,7,\ldots .$$

The quantity

$$R_h = \frac{Q_h - Q_{h+1}}{e_h} = \frac{Q_h - Q_{h+1}}{Q_h} H_{hn}, \; h=0,1,2,\ldots$$

is approximately N(0,1) for large samples (O. Anderson, 1929). Hence if an h_o is found such that $R_{h_o}-1$ is significant but R_{h_o} is not significant at the chosen significance level it is assumed that the systematic part of the series has been approximately eliminated in the h_oth difference series. It should be pointed out that the choice of the order of difference h_o is a multiple choice problem. Therefore the maintaining of a fixed level of significance is extremely difficult.

The estimate Q_{h_o} of σ^2 is not completely efficient since one observation is lost each time one takes a higher difference. Morse and Grubbs (1947) have treated this problem and present a table for evaluating the efficiency for various n and h_o. The higher the order of the series of differences from which the variance has been estimated the less efficient is the estimate.

Applicability of the variate difference method is thought to be limited even by Tintner whose work with this method has been extensive. This method is not valid when the errors are autocorrelated and an autoregressive scheme should be used under such circumstances.

Oscillatory and Periodic Movements

In some types of data a distinct oscillatory movement may be apparent. Suppose it is known that such a time series x_t, $t=\ldots, -1,0,+1,\ldots$, has the periodic parametric form

$$x_t=A_o+\sum_{p=1}^{k}\left[A_p\cos\omega_p t+B_p\sin\omega_p t\right]+\epsilon_t$$

where ϵ_t is a random element with mean zero and variance σ^2, A_o, A_p, B_p, and ω_p are known real constants and $0\leq\omega_p\leq\pi$. For simplicity, it is assumed for the moment that A_o is zero.

Consider a sample $(x_0,\ldots, x_{n-1})$ from the time series. If one multiplies through by $\cos\omega t$, $0\leq\omega\leq\pi$, sums over t and divides by n then

$$\begin{aligned}\alpha(\omega)&=\frac{1}{n}\sum_{t=0}^{n-1} x_t\cos\omega t\\ &=\frac{1}{n}\sum_{t=0}^{n-1}\left[\sum_{p=1}^{k}(A_p\cos\omega_p t\ \cos\omega t+B_p\sin\omega_p t\ \cos\omega t)+\epsilon_t\cos\omega t\right].\end{aligned}$$

The expression can now be rewritten

$$\alpha(\omega)=\frac{1}{n}\sum_{p=1}^{k}\left[\frac{A_p}{2}\sum_{t=0}^{n-1}\cos(\omega_p+\omega)t+\cos(\omega_p-\omega)t\right.$$

$$\left.+\frac{B_p}{2}\sum_{t=0}^{n-1}\sin(\omega_p+\omega)t+\sin(\omega_p-\omega)t\right]$$

$$+\frac{1}{n}\sum_{t=0}^{n-1}\epsilon_t\cos\omega t.$$

Now as $n\longrightarrow\infty$, $\alpha(\omega)\longrightarrow A_p/2$, $\omega=\omega_p$, $p=1,\ldots,k$

$$\longrightarrow 0, \omega\neq\omega_p.$$

Similarly, if the original expression is multiplied through by $\sin\omega t$ and denoted by $\alpha'(\omega)$, and the corresponding operations carried out, as $n\longrightarrow\infty$,

$$\alpha'(\omega)\longrightarrow -B_p/2, \omega=\omega_p,\ p=1,\ldots,k$$

$$\longrightarrow\ 0,\ \omega\neq\omega_p.$$

Therefore, if $2\pi/\omega_p$ is a genuine period of the time series x_t, $t=\ldots,-1,0,+1,\ldots$, $\frac{1}{n}\sum_{t=0}^{n-1}x_t\cos\omega_p t$ will tend to be near $A_p/2$ and $\frac{1}{n}\sum_{t=0}^{n-1}x_t\sin\omega_p t$ will tend to be near $-B_p/2$.

Let n in the sample $(x_1,\ldots, x_n)$ be odd, say $2r+1$, and let

$$p=\frac{2\pi p}{2r+1},\ p=1,\ldots,k,\quad r\geqslant k.$$

The form of the periodic function is now

$$x_t=A_o+\sum_{p=1}^{k}\left[A_p\cos p\theta_t+B_p\sin p\theta_t\right]+\epsilon_t$$

where $\theta_t=2\pi t/2r+1$.

If the periods of the time series are known but the A_o, A_p, B_p, $p=1,\ldots,k$ are unknown, the minimum variance estimators for these quantities can be found by least squares. These least squares equations are

$$\sum_{t=0}^{n-1}\left[x_t-A_o-\sum_{p=1}^{k}(A_p\cos p\Theta_t+B_p\sin p\Theta_t)\right]=0$$

and

$$\sum_{t=0}^{n-1}\left[x_t-A_o-\sum_{p=1}^{k}(A_p\cos p\Theta_t+B_p\sin p\Theta_t)\right]\begin{bmatrix}\sin h\Theta_t\\ \cos j\Theta_t\end{bmatrix}=0,$$

$$h,\ j=1,\ldots,k.$$

The standard formula for the sum of a cosine progression gives

$$\sum_{t=0}^{n-1}\cos mt\phi=\sin(1/2)mt\phi\ \cos(1/2)m(t-1)\phi/\sin(1/2)m\phi$$

where $\phi=\Theta_t/t$. For integral values of m this will vanish since $(1/2)mt\phi=m\pi$. Therefore all sums of the form

$$\sum_t\cos h\Theta_t\ \cos j\Theta_t=1/2\left[\sum_t\cos(h+j)\Theta_t+\sum_t\cos(h-j)\Theta_t\right]$$

will vanish unless $h=j$, when the value will be $(1/2)n$.

The first of the least squares equations gives the result $A_o=\sum_{t=0}^{n-1}x_t/n=\bar{x}$.

From the normal equations, the covariance matrix is found to be

$$\begin{Vmatrix} n/2 & 0 & . & . & . & 0 \\ 0 & n/2 & . & . & . & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & . & . & . & n/2 \end{Vmatrix}$$

and its inverse is

$$\begin{Vmatrix} 2/n & 0 & . & . & . & 0 \\ 0 & 2/n & . & . & . & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & . & . & . & 2/n \end{Vmatrix}$$

It follows that coefficients are random variables with zero covariances and variances $2\sigma^2/2r+1$. The variance of a fitted value is $(2k+1)\sigma^2/n$ and is independent of the angle θ_t. The residual sum of squares $\sum_{t=0}^{n-1}(x_t-\bar{x}_t)^2$ is given by

$$\sum_{t=0}^{n-1} x_t - na_o^2 - 1/2\left[n\sum_{p=1}^{k}(a_p^2+b_p^2)\right].$$

The expectation of this is $(n-2k-1)\sigma^2$ and so the variance of an observation is estimated by

$$s^2 = \frac{\sum_{t=0}^{n-1}(x_t-\bar{x}_t)^2}{n-2k-1}.$$

If the time series has periodic form but the A_p, B_p, and ω_p and even k are unknown, a method of searching for suspected periods is necessary. The behavior of the mean values of $\alpha(\omega)$ and $\alpha'(\omega)$ described earlier suggests that these values considered as functions of ω might be useful in screening out true periods if any

exist.

As early as 1898, Schuster proposed a method of searching for possible periods. Walker in 1914, and Fisher in 1929, followed with methods of testing the significance of suspected periods. This method, generally known as periodogram analysis, tests the significance of possible periods under the assumption that x_t is a white noise.

If it is assumed that there are no true periods at all in the given time series, then the A_p and B_p are all zero. But if $2\pi/\omega_p$ is a true period, the behavior of $\alpha(\omega_p)$ and $\alpha'(\omega_p)$ indicate that both of these will tend to have values away from zero for large n. The quantity $\alpha^2(\omega_p)+\alpha'^2(\omega_p)$ has a value for each ω_p, p=1,...,k, so one needs a way of testing whether the largest (or mth largest) of these quantities is significantly large under the assumption that x_t is a white noise.

The further assumption that x_t is normal white noise with variance σ^2 is made in dealing with the problem of significance testing. The quantities

$$\sqrt{\frac{2(2r+1)}{\sigma^2}}\,\alpha(\omega_p),\ \sqrt{\frac{2(2r+1)}{\sigma^2}}\,\alpha'(\omega_p),\ p=1,\ldots,k$$

are 2k independent random variables distributed N(0,1). If

$$u_p=\frac{2k+1}{\sigma^2}\left[\alpha^2(\omega_p)+\alpha'^2(\omega_p)\right],\ p=1,\ldots,k,$$

$u_1,\ldots,u_k$ are chi square variables with 2 degrees of freedom, and have probability element

$$e^{-(u_1+\ldots+u_k)}du_1\ldots du_k$$

for $u_p \geq 0$, $p=1,\ldots,k$, and 0 otherwise. The problem of whether the largest (or mth largest) of the quantities $\alpha^2(\omega_p)+\alpha'^2(\omega_p)$ is significantly large reduces to testing whether the largest of $u_1,\ldots, u_k$ is significantly large. The test which suggests itself, since σ^2 is unknown, is whether the largest (or mth largest) of the ratios

$$g = \frac{u_p}{u_1+\ldots+u_k},\ p=1,\ldots,k$$

is significantly large.

Walker's (1914) criterion was that the chance for the largest intensity to exceed a given level x is given by $1-(1-e^{x/2})^k$. Fisher (1929) found the distribution function of g:

$$P(g > g')=\sum_{p=0}^{r}(-1)^p\binom{k}{p+1}\left[1-(p+1)g'\right]^{k-1}$$

where r is the largest integer $\leq k-1$ for which $1-(r+1)g' \geq 0$. For a given k and a given α the value of g_α for which $P(g > g_\alpha)=\alpha$ would be the critical value of g for significance level 100α %. Tabulations of g_α have been made by Davis (1941) for a wide range of values of g_α and k.

Similarly if g is defined as the mth largest of $u_1,\ldots, u_k$ and divided by $u_1+\ldots+u_k$, then

$$P(g > g')= \frac{k!}{(m-1)!}\sum_{p=m}^{r}\frac{(-1)^{p-m}(1-pg')^{k-1}}{p(k-p)!(p-m)!}$$

where r is the largest integer for which $1-rg' \geq 0$.

Hartley (1949) proposed a method for testing the significance of periods using the F ratio. The observed intensities $S=a_p^2+b_p^2$

are computed and the significance of the largest intensity is tested.

Hartley starts with the hypothesis of a completely random series where the x_t are normal deviates with variance σ^2. The p intensities $1/2\left[nS_p^2\right]$, $p=1,\ldots,k$, are all independent χ^2 variates, each with 2 degrees of freedom. Walker's criterion is converted into an exact test by making use of the residual as an independent estimate of σ^2. The test that results is one for the maximum variance ratio

$$F_{max}=1/4\left[nS_{max}^2(n-2k-1)/R^2\right].$$

The probability for $F_{max}\leq F^*$ is given by

$$P(F^*)=\int_0^{\infty}\phi_{\nu}(s)(1-\exp\left[-s^2F^*\right])^k ds$$

where $\phi_{\nu}(s)$ denotes the distribution of a sample standard deviation based on ν degrees of freedom. For this problem, $\nu=n-2k-1$. Hartley uses an approximation to the integral valid only for upper percentage points. Instead of evaluating the upper 100α % point of the distribution, the $100\alpha/k$ % point of the F distribution based on 2 and ν degrees of freedom.

If the series x_t is of the periodic form, then of the k periods examined, some, say h, have positive amplitudes and the remaining k-h have zero amplitudes. This says that $A_p^2+B_p^2>0$ for h values of p, and $A_p^2+B_p^2=0$ for k-h values of p. If the maximum observed intensity is judged significant by the test, the conclusion that $A_p^2+B_p^2>0$ for that particular p for which the maximum intensity

was observed is not strictly valid. Only rejection of the hypothesis of randomness is justified. In practice, however, it is usually desired to conclude that the maximum intensity observed indicates that for the particular p the $A_p^2+B_p^2>0$.

To investigate the extent to which the experimenter might be misled by the significant test, it is necessary to investigate the power of the test. The chance of reaching a significant result is the sum of two situations:

1. The observed maximum intensity $1/2[nS_p^2]$ does come from the set of h positive intensities, and
2. The observed maximum intensity does come from the set of k-h true zero intensities.

A wrong conclusion would be reached if the second situation occurs. It has been shown by Hartley (1949) that this chance is smaller than (k-h)/k times the error of the first kind, and hence, if the independent harmonic intensities are used in the F_{max} test, the chance of reaching a wrong conclusion is almost negligible.

In order to examine the F_{max} test under the general hypothesis

$$H_M: x_t=M_t+\epsilon_t,\ t=0,1,\ldots,n-1,$$

where the ϵ_t are random normal deviates, the x_t are represented by their complete, finite Fourier expansion with n assumed odd for convenience. The general hypothesis can now be written

$${}_{1/2(n-1)}H_h: x_t=A_o+\sum_{p=1}^{1/2(n-1)}(A_p\cos p\theta_t+B_p\sin p\theta_t)+\epsilon_t.$$

This differs from the previously stated ${}_kH_h$ in that the representation of x_t includes (n-1)-k additional real Fourier terms. The

magnitude of these additional terms as a percentage of the variance of the ϵ_t can be expressed by a non-centrality ratio (Hartley, 1949)

$$\delta = (1/2)n \sum_{k+1}^{1/2(n-1)} (A_p^2+B_p^2)/\sigma^2.$$

If the series of x_t are the ordinates of a smooth function, then from standard Fourier theory it is known that for a sufficiently large k, δ can be made as small as is required for any $n > 2k$. In practice then, if periods up to order m are suspected, the F_{max} test will detect only these if $n \gg 2m$. If too small values of m and n are used, the F_{max} test, which is based on the assumption that δ is zero, will be biased by an amount depending on the value of δ. This effect can be calculated exactly using the methods of Hartley. It is stressed that the F_{max} test is inappropriate unless the x_t can be represented by a moderate number of Fourier terms and yet δ will be expected to be zero or small.

The residuals from the final fitted curve can be examined for autocorrelation to see if the periodic behavior has been adequately described. Use of the circular coefficient of autocorrelation with lag L is appropriate in this case. For lag 1, Anderson and Anderson (1950) have calculated tables of significance points of R where

$$R = \frac{\sum_{t=1}^{n}(x_t-M_t)(x_{t-1}-M_{t-1})}{\sum_{t=1}^{n}(x_t-M_t)^2}, \quad x_0=x_n$$

is the circular autocorrelation coefficient used for residuals

from a Fourier series. Since economic data were the primary motivation for the work, the basic periods for which the R_α values were calculated are for p=2, 3, 4, 6 and 12, indicating the usual yearly increments for which economic data are tabulated. Significance levels are for α=.01 and α=.05 and N ranges from 6 to the point in each distribution where tables for the regular coefficient of correlation or of the incomplete beta function give satisfactory approximations.

Using the same data for which the correlogram was plotted in Fig. 2, harmonic analysis and Hartley's method can be used to test for periods in the data. If it is assumed that the periods will be no shorter than one month in length, then an upper limit of 12 can be used in searching for periods in the data. The results of such a harmonic analysis are given in Table 2.

The Hartley test at the 1% level of significance gives periods 1, 2, 3, and 4 as significant. Therefore the resulting model is

$$x_t=A_o+\sum_{p=1}^{4}(A_p\cos\frac{2\pi pt}{365}+B_p\sin\frac{2\pi pt}{365}).$$

The coefficient of determination resulting from this model is .6353 and multiple R will be .7971. The standard deviation of an observation is .0425.

To see whether the systematic portion of the variation has been sufficiently explained, the autocorrelation of the residuals is examined. Table 3 gives the coefficients of autocorrelation for the residuals from the regression line. None of the coefficients calculated are near to exceeding the significance level

for a reasonable α. It seems valid to conclude then that the data are periodic in nature and that the fitted model gives good estimates of the true parameters.

Table 2. Harmonic Analysis of Probabilities of Dry Days for Manhattan, Kansas.

a_o = .849454				
p	a	b	Intensity	F Ratio
1	.073138	-.009187	.005433	279.53*
2	.005515	.017995	.000354	18.22*
3	-.008014	-.012171	.000121	10.93*
4	.005011	.010217	.000129	6.66*
5	-.000671	-.001261	.000002	0.10
6	-.002125	.006155	.000042	2.18
7	-.002107	-.001613	.000007	0.36
8	-.002280	-.003540	.000018	0.91
9	.004523	-.002480	.000027	1.37
10	-.006448	-.004420	.000061	3.14
11	-.000200	.002886	.000008	0.43
12	-.004532	-.005387	.000050	2.55

* significant at 1% level

Table 3. Coefficients of Autocorrelation of Residuals from Regression Model with p=4.

Lag	R_L	Lag	R_L	Lag	R_L
1	.0979	8	-.0602	15	-.0169
2	-.0544	9	-.0785	16	-.0949
3	-.0771	10	.0146	17	-.0216
4	.0524	11	-.0482	18	-.0746
5	-.0004	12	.0345	19	-.0690
6	-.0190	13	.0167	20	-.0038
7	-.1091	14	.0183		

SUMMARY AND CONCLUSIONS

Problems involved in the analysis of time series data have concerned statisticians since statistics emerged as a separate discipline. Most of the well-known statisticians have, at one time or another, made some contribution to the theory of time series analysis.

The first area of the problem that was attacked was that of testing for the existence of autocorrelation, and considerable progress has been made. Less well developed are the areas of estimation and hypothesis testing. It has been illustrated here that significant contributions are still being made and much remains to be done in these areas. Small sample theory is extremely vague and efficient goodness-of-fit tests are practically non-existent.

Were it not for the fact that most time series, particularly those in economics, are relatively short, the non-independence of the ϵ's would pose a much less serious problem. The theory of least squares estimation should be used wherever applicable. It is possible that fitting the ordinary least squares line is the best starting point in the analysis of a time series.

The methods presented here are generally amenable to programming for computers.

ACKNOWLEDGMENT

The writer wishes to express gratitude to Dr. A. M. Feyerherm for his invaluable assistance in the preparation of this report. Deeply appreciated are the many helpful suggestions provided by him during the entire period of graduate study.

REFERENCES

Anderson, O.
Die korrelationsrechnung in der konjunkturforschung. Frankfurt: 1929.

Anderson, R. L.
Distribution of the serial correlation coefficient. Annals of Mathematical Statistics. 13:1-13. 1942.

________.
The problem of autocorrelation in regression analysis. Journal of the American Statistical Association. 49:113-29. 1954.

Anderson, R. L., and T. W. Anderson.
Distribution of the circular serial correlation coefficient for residuals from a fitted Fourier Series. Annals of Mathematical Statistics. 21:59-81. 1950.

Anderson, T. W.
On the theory of testing serial correlation. Skandinavisk Aktuarietidskrift. 31:88-116. 1948.

Bartlett, M. S.
On the theoretical specification and sampling properties of autocorrelated time series. Journal of the Royal Statistical Society, Series B. 8:34-35. 1946.

Champernowne, D. G.
Sampling theory applied to autoregressive sequences. Journal of the Royal Statistical Society, Series B. 10:204-42. 1948.

Cochrane, D., and G. H. Orcutt.
Application of least-squares regression to relationships containing autocorrelated error terms. Journal of the American Statistical Association. 44:32-61. 1949.

Davis, H. T.
The analysis of economic time series. Bloomington, Indiana: Principia Press, 1941.

Durbin, J.
Estimation of parameters in time-series regression models. Journal of the Royal Statistical Society, Series B. 22:139-53. 1960.

Durbin, J., and G. S. Watson.
Testing for serial correlation in least-squares regression. Biometrika. 38:159-78. 1951.

Ezekiel, Mordecai, and Karl A. Fox.
Methods of correlation and regression analysis. New York: John Wiley and Sons, 1959.

Fisher, R. A.
Tests of significance in harmonic analysis. Proceedings of the Royal Society of London, Series A. 125:54-59. 1929.

Fuller, Wayne A., and James E. Martin.
The effects of autocorrelated errors on the statistical estimation of distributed lag models. Journal of Farm Economics. 43:71-82. 1961.

Guest, P. G.
Numerical methods of curve fitting. London: Cambridge University Press, 1961.

Hannan, E. J.
Time series analysis. New York: John Wiley and Sons, 1960.

Hart, B. I.
Significance levels for the ratio of the mean square successive difference to the variance. Annals of Mathematical Statistics. 13:445-47. 1942.

Hartley, H. O.
Tests of significance in harmonic analysis. Biometrika. 36: 194-201. 1949.

________.
The Modified Gauss-Newton Method for the fitting of non-linear regression functions by least-squares. Technometrics. 3:269-80. 1961.

Hurwicz, L.
Least squares bias in time series. Statistical inference in dynamic economic models, T. C. Koopmans, editor. New York: John Wiley and Sons, 1950.

Koopmans, T. C.
Serial correlation and quadratic forms in normal variables. Annals of Mathematical Statistics. 13:14-23. 1942.

Mann, H. B., and A. Wald.
On the statistical treatment of linear stochastic difference equations. Econometrica. 11:173-220. 1943.

Moran, P. A. P.
A test for the serial independence of residuals. Biometrika. 37:178-81. 1950.

Morse, A. P., and F. E. Grubbs.
The estimation of dispersion from differences. Annals of Mathematical Statistics. 18:194-214. 1947.

Schuster, A.
On the investigation of hidden periodicities with application to a supposed 26-day period of meterological phenomena. Terrestrial Magnetism. 3:13-41. 1898.

Tintner, Gerhard.
The variate difference method. Bloomington, Indiana: Principia Press, 1940.

_______.
Econometrics. New York: John Wiley and Sons, 1952.

von Neumann, J.
Distribution of the ratio of the mean square successive difference to the variance. Annals of Mathematical Statistics. 12:367-95. 1941.

Wald, A., and J. Wolfowitz.
An exact test for randomness in the non-parametric case based on serial correlation. Annals of Mathematical Statistics. 14:378-88. 1945.

Walker, Gilbert.
On the criterion for the reality of relationships or periodicities. Calcutta Indian Meterological Memoirs. 21. 1914.

Watson, G. S.
Serial correlation in regression analysis. Unpublished thesis, North Carolina State College, Raleigh, North Carolina, 1951.

Wilks, S. S.
Mathematical statistics. New York: John Wiley and Sons, 1962.

Wold, Herman.
A study in the analysis of stationary time series. Second edition. Stockholm: Almqvist and Wiksell, 1954.

Yule, G. U.
Why do we sometimes get non-sense correlations between time series?. Journal of the Royal Statistical Society. 89:1-64. 1926.

APPENDIX

Probability of a Dry Day* at Manhattan, Kansas

Day	Probability	Day	Probability	Day	Probability
1	.[illegible]	41	.[illegible]	81	.86206896
2	.93103448	42	.94827586	82	.91379348
3	.89655172	43	.94827586	83	.82758672
4	.84482758	44	.81034482	84	.84482758
5	.91379310	45	.89655177	85	.87931010
6	.93103448	46	.91379310	86	.94827548
7	.93103448	47	.86206896	87	.81034448
8	.93103448	48	.89655172	88	.86206848
9	.89655172	49	.84482758	89	.84482772
10	.93103448	50	.93103479	90	.89655148
11	.94827586	51	.86206896	91	.86206806
12	.94827586	52	.91379310	92	.91379386
13	.93103448	53	.91379310	93	.79310348
14	.96551724	54	.86206896	94	.82758624
15	.93103448	55	.89655179	95	.84482748
16	.93103448	56	.93103448	96	.91379348
17	.91379310	57	.86206896	97	.81034410
18	.94827586	58	.82758620	98	.75862086
19	.93103448	59	.84482758	99	.81034448
20	.96551724	60	.94827586	100	.79310342
21	.89655172	61	.87931044	101	.79310372
22	.96275862	62	.86206896	102	.82758662
23	.86206896	63	.86206896	103	.86206896
24	.94827586	64	.82758620	104	.81034486
25	.94827586	65	.87931044	105	.79310386
26	.93103448	66	.91379310	106	.81034448
27	.89655172	67	.94827586	107	.87931072
28	.96551724	68	.93103476	108	.84482724
29	.93103448	69	.87931034	109	.77586248
30	.91379310	70	.86206896	110	.77586210
31	.94827586	71	.84482758	111	.79310386
32	.96551724	72	.87931034	112	.84482724
33	.89655172	73	.89655172	113	.82758672
34	.84482758	74	.86206896	114	.79310358
35	.91379310	75	.87931034	115	.74137910
36	.94827586	76	.94827586	116	.82758686
37	.96551724	77	.82758620	117	.75862024
38	.91379310	78	.81034482	118	.81034410
39	.84482758	79	.91379310	119	.79310358
40	.94827586	80	.93103448	120	.77586286

*Dry day is defined as less than .10" precipitation. Probabilities are based on data for years 1901-1960.

APPENDIX (continued)

Day	Probability	Day	Probability	Day	Probability
121	.79310344	171	.74137944	221	.79310344
122	.77586206	172	.79310344	222	.84482706
123	[illegible]	173	.81034482	223	.77586258
124	.82758620	174	.77586206	224	.81034420
125	.74137931	175	.79310344	225	.82758631
126	.72413793	176	.75862068	226	.75862093
127	.77586206	177	.79310344	227	.75862006
128	.79310344	178	.77586206	228	.77586244
129	.77586206	179	.75862068	229	.81034406
130	.72413793	180	.77586206	230	.84482793
131	.79310344	181	.74137931	231	.81034444
132	.81034482	182	.84482778	232	.75862082
133	.75862068	183	.72413796	233	.86206868
134	.79310344	184	.79310344	234	.86206844
135	.84482758	185	.77586206	235	.82758658
136	.74137931	186	.79310344	236	.82758631
137	.72413793	187	.75862068	237	.87931093
138	.75862068	188	.77586206	238	.75862068
139	.68965517	189	.81034482	239	.75862017
140	.84482758	190	.86206896	240	.77586258
141	.74137931	191	.84482758	241	.91379331
142	.77586206	192	.82758620	242	.89655106
143	.68965517	193	.81034496	243	.86206817
144	.72413793	194	.81034462	244	.79310393
145	.75862068	195	.79310344	245	.81034468
146	.77586206	196	.84482758	246	.81034406
147	.72413793	197	.77586206	247	.81034493
148	.67241379	198	.81034482	248	.87931079
149	.75862068	199	.82758620	249	.75862066
150	.79310344	200	.79310344	250	.84482744
151	.70689655	201	.84482758	251	.82758655
152	.77586206	202	.91379350	252	.82758606
153	.72413793	203	.87931034	253	.75862093
154	.79310344	204	.87931034	254	.81034444
155	.82758620	205	.87931047	255	.68965520
156	.65517241	206	.87931034	256	.82758641
157	.72413793	207	.82758620	257	.77586293
158	.72413793	208	.79310344	258	.77586293
159	.77586206	209	.87931044	259	.79310306
160	.70689655	210	.86206896	260	.84482755
161	.67241379	211	.75862068	261	.87931079
162	.75862068	212	.79310344	262	.91379368
163	.72413793	213	.87931034	263	.91379393
164	.75862068	214	.75862068	264	.87931068
165	.82758620	215	.86206896	265	.82758620
166	.74137931	216	.91379310	266	.93103431
167	.77586206	217	.86206896	267	.82758606
168	.84482758	218	.81034496	268	.86206858
169	.75862068	219	.70689675	269	.74137968
170	.77586206	220	.79310344	270	.79310306

APPENDIX (continued)

Day	Probability
271	.81034482
272	.81034482
273	.86206896
274	.86206896
275	.91379310
276	.84482758
277	.86206896
278	.91379310
279	.81034482
280	.86206896
281	.81034482
282	.81034482
283	.91379310
284	.89655172
285	.82758620
286	.87931034
287	.81034482
288	.86206896
289	.89655172
290	.91379310
291	.82758620
292	.87931034
293	.86206896
294	.91379310
295	.82758620
296	.91379310
297	.87931034
298	.91379310
299	.86206896
300	.91379310
301	.89655172
302	.87931034
303	.82758620
304	.81034482
305	.89655172
306	.87931034
307	.84482758
308	.86206896
309	.93103448
310	.89655172
311	.91379310
312	.91379310
313	.87931034
314	.94827586
315	.93103448
316	.93103448
317	.91379310
318	.87931034
319	.91379310
320	.87931034

Day	Probability
321	.91379310
322	.94827586
323	.89655172
324	.89655172
325	.89655172
326	.91379310
327	.96551724
328	.96551724
329	.87931034
330	.86206896
331	.89655172
332	.87931034
333	.84482758
334	.91379310
335	.89655172
336	.91379310
337	.87931034
338	.89655172
339	.75862068
340	.89655172
341	.98275862
342	.98275862
343	.86206896
344	.93103448
345	.91379310
346	.89655172
347	.91379310
348	.91379310
349	.96551724
350	1.00000000
351	.91379310
352	.96551724
353	.94827586
354	.91379310
355	.96551724
356	.89655172
357	.84482758
358	.86206896
359	.94827586
360	.87931034
361	.89655172
362	.93103448
363	.98275862
364	.98275862
365	.87931034

*   *   *   *

ANALYSIS OF STATIONARY TIME SERIES

by

GAIL EUGENE BACHMAN

B. A., University of Wichita, 1959

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1963

In the sciences many problems occur in which a process produces what may be considered a family of random variables such that there is a value of X_t for each value of t in some interval T. The experimenter usually wishes to investigate the nature of the response curve over the interval T.

The usual regression model in least squares analysis is

$$X_t=\beta_0+\sum_{i=1}^{r}\beta_i Z_{it}+\epsilon_t,\ t=0,1,\ldots,n-1,$$

where the Z's are assumed fixed in repeated sampling and the ϵ's are independently distributed with mean zero and variance σ^2. In applying traditional least squares methods to time series data difficulties may arise because successive observations often lack the property of independence. If the ϵ's are not independent the assumptions necessary for using ordinary least squares estimation theory are violated.

A concern for problems involved in analyzing time series data was lacking among research workers until the middle 1920's. In the 1930's Koopmans, Wold, and others clarified the sampling significance of regression analysis of time series data. Considerable work followed on the problem of testing for the existence of correlation of the errors but all too little on the more important problem of the best estimation procedure when correlations do exist. Results are still somewhat lacking in this area, but several estimation procedures have been proposed since 1950. This paper deals with the methods generally used in the social and biological sciences.

The discussion of time series is usually confined to what are called stationary time series. Roughly, this means that the time series is without trends, not only in the mean values of the X_t but also in their variances.

Three models are generally used in analyzing time series data. If it is assumed that the data follow an underlying systematic scheme with random fluctuations superimposed, the methods of harmonic analysis and periodogram analysis may be used to determine the nature of a systematic component with regular periods. For a systematic component with an irregular oscillatory movement the variables Z_{it} in the regular regression model may be replaced by t^i's and a polynomial form used to locally describe the function. In other cases, the assumed model may involve lagged values of X as predictors or possibly both lagged X's and Z's. This is referred to as an autoregressive model.

Much remains to be done on the time series problem in the areas of estimation and hypothesis testing. Small sample theory is extremely vague and efficient goodness of fit tests are practically non-existent.